Урок 12. Как компьютер распознаёт речь и синтезирует голос?

5 класс · Урок 12 из 17 · 45 минут

Вся программа 5 класса · Технологии искусственного интеллекта

Чем различаются распознавание и синтез речи. Ученики проведут опыт с расшифровкой, проверят озвучивание музейной подписи и разберут необычную просьбу знакомым голосом.

Содержание урока

- 01. Цель урока
- 02. Ход урока
- 03. Этап 1. Одна запись — два места встречи · 5 минут
- 04. Этап 2. Звук и текст: разные входы и результаты · 10 минут
- 05. Этап 3. Когда звук не даёт готового ответа · 11 минут
- 06. Этап 4. Озвучивание музейной подписи · 12 минут
- 07. Этап 5. Знакомое звучание и проверка отправителя · 7 минут

01. Цель урока

Понять, как компьютер преобразует речь в текст и текст в звук, и научиться проверять результат и собеседника в голосовом общении.

К концу урока ученики смогут:

- различать распознавание и синтез речи по входным данным и результату;
- объяснять роль обучающих записей и причины ошибок на новых голосах и в шуме;
- различать омофоны, омографы и омонимы на примерах;
- сверять расшифровку с исходной речью и проверять произношение и смысл озвученного текста;
- называть области применения речевых технологий и способ проверки необычной просьбы знакомым голосом.

02. Ход урока

Этап	Мин	Учитель	Ученики
1. Одна запись — два места встречи	5	Вводит тему через ошибочную голосовую заметку	Выбирают источник для проверки
2. Звук и текст: разные входы и результаты	10	Объясняет обучение и работу речевых технологий	Читают схему, сравнивают человека и компьютер
3. Когда звук не даёт готового ответа	11	Объясняет неоднозначность и организует опыт	Записывают фразы, дополняют контекстом и проверяют вывод
4. Озвучивание музейной подписи	12	Организует синтез и проверку по смысловым условиям	Редактируют подпись, слушают результат и записывают наблюдение
5. Знакомое звучание и проверка отправителя	7	Объясняет работу голосового бота и проводит итоговую проверку	Решают, как проверить просьбу, отвечают на вопросы
Итого	45		

03. Этап 1. Одна запись — два места встречи

5 минут

ГОВОРИМ УЧЕНИКАМ

Сегодня выясним, как компьютер записывает нашу речь словами и как озвучивает написанное. Это разные задачи, хотя обе связаны с голосом. Для одной нужен исходный звук, для другой — текст. Мы проверим обе технологии и разберёмся, почему узнаваемый голос ещё не доказывает личность отправителя.

Где собираться?

Организатор диктует голосовую заметку о встрече кружка на шумной улице. В текстовой версии появляется: «Собираемся в парке». Школьнику показалось, что речь шла об арке у входа в парк. Как выяснить место встречи?

Проверьте основание

Предложите два способа выяснить, что сказал организатор. Затем послушайте исходную фразу от учителя и запишите верный вариант. Объясните, почему одного предположения о привычном месте встречи недостаточно.

Для учителя

Читайте: «Собираемся у арки». Можно переслушать исходную запись или уточнить у организатора по известному контакту. Угадывать по прошлым встречам недостаточно.

Случай вымышленный: шум скрыл начало сочетания «у арки», а оставшегося звучания не хватило для надёжного различения с «в парке». Ответ реального сервиса здесь не проверяли. Работающий распознаватель в опыте не участвует. Чтение учителя сообщает классу, что сказал организатор.

ГОВОРИМ УЧЕНИКАМ

В нашем сценарии организатор назвал арку, а в расшифровке оказался парк. Шум скрыл часть звуков, нужных для различения этих мест. Учитель сообщил исходную фразу, и мы смогли сравнить её с текстом. Если настоящая запись неразборчива, место нужно уточнить у организатора, а не выбирать по памяти.

04. Этап 2. Звук и текст: разные входы и результаты 10 минут

ГОВОРИМ УЧЕНИКАМ

Распознавание речи — получение текста из звучащего сообщения. Микрофон улавливает звук, а компьютер хранит его как числовые данные. Модель обрабатывает эту запись и выбирает слова, которые могли прозвучать. Если убрать микрофон и запись, у распознавателя не будет исходного звука для работы.

Распознавание часто обучают на записях речи вместе с правильными текстами. Запись показывает звучание, а текст сообщает, какие слова были сказаны. По таким парам модель учится связывать особенности звука со словами. Если в новых условиях появляются непривычная речь, быстрый темп или шум, ошибки могут стать чаще.

Человек тоже использует слух и может не расслышать сообщение. Но собеседник видит обстановку, замечает жесты, помнит общий разговор и может сразу переспросить. Программа получает только переданные ей данные. Нужно знать, получила ли программа только звук или ещё и сведения о разговоре.

Синтез речи начинает с текста и создаёт звуковой сигнал. По примерам текста и соответствующей речи модель учится произношению. При озвучивании она выбирает звуки, ударения и паузы. Динамик воспроизводит созданный сигнал; сам динамик не решает, что сказать.



Распознавание записывает сказанное. Синтез озвучивает написанное.

ГОВОРИМ УЧЕНИКАМ

Распознавание нужно для субтитров, голосового поиска и записи интервью. Синтез помогает слушать электронный текст и голосовые подсказки. Разговорный помощник ещё должен обработать запрос и подготовить ответ. Одна запись слов ещё не даёт ответа на вопрос.

Реальный случай: приложение учится особенностям речи

ГОВОРИМ УЧЕНИКАМ

9 ноября 2021 года Google пригласила первых участников из пяти стран тестировать Project Relate — приложение для людей, чью речь обычные системы распознают с трудом. Участники записывали фразы, чтобы приложение лучше учитывало особенности их речи. Режим Listen превращал сказанное в текст, а Repeat повторял сообщение синтезированным голосом. Здесь технологии помогали общению и выполняли разные части работы.

05. Этап 3. Когда звук не даёт готового ответа 11 минут

ГОВОРИМ УЧЕНИКАМ

Некоторые слова звучат одинаково, хотя пишутся по-разному. Например, «гриб» и «грипп» при обычном произношении звучат одинаково: это омофоны. В конце слова звонкий звук может произноситься как глухой, поэтому по одному звучанию не всегда видно написание. Контекст помогает выбрать слово.

Есть и другая трудность. «А́тлас» с картами и «атла́с» как ткань пишутся одинаково, но различаются ударением: это омографы. Если ударение хорошо слышно, оно уже позволяет различить эти слова. При синтезе текста без знака ударения выбрать произношение помогает контекст. А «коса» как причёска и «коса» как инструмент одинаковы и по звучанию, и по написанию, но имеют разные значения: это омонимы. Распознавателю и

синтезатору приходится решать разные стороны этих задач: одному выбирать запись по звуку, другому — звучание по записи.

Задание 1. Что изменил контекст?

Учитель произнесёт короткую фразу, а затем более полное сообщение. Не смотрите в его ответ заранее.

1. Запишите короткую фразу. Если возможны разные слова, отметьте варианты.
2. После полного сообщения внесите исправление или сохраните первую запись. Подчеркните сведения, на которых основано решение.
3. По второму сообщению объясните, что можно определить по контексту, даже если часть звучания потеряна.
4. Предложите свою фразу с неоднозначным словом. Сосед задаёт один уточняющий вопрос.

Результат: первая запись, решение после контекста и собственный вопрос. Обоснуйте решение сведениями из сообщения.

Для учителя

Первое чтение: «На экране появился код». Читайте естественно, без подчёркивания последнего согласного. В обычной речи «код» и «кот» звучат одинаково: в конце слова «д» оглушается. На экране могут появиться и цифры, и изображение животного. Второе чтение: «На экране появился код: четыре цифры для входа в игру». Продолжение позволяет выбрать написание «код». Верная первая запись без объяснения ещё не показывает, что неоднозначность замечена.

Второе сообщение: «В атласе показаны острова и морские течения». При чтении произносите «в а́тласе». Даже если ударение не расслышано, слова «острова» и «морские течения» указывают на книгу с картами. Основание нужно и для верной первой записи.

Собственная фраза подходит, если вопрос соседа помогает различить значения. Например, «На фотографии видна коса» — «Что ещё видно: голову человека или рукоятку инструмента?» Не требуйте именно этого примера.

Это опыт человеческого слушания и роли контекста, а не замер точности компьютерной модели. Он показывает, какие сведения доступны для решения; человек и программа не устроены одинаково. При нарушении слуха дайте две письменные версии с пропуском спорного слова.

ГОВОРИМ УЧЕНИКАМ

На экране могли появиться и кот, и код. Одинакового звучания не хватало для выбора написания. Сведения о четырёх цифрах для входа позволили записать «код». Во втором сообщении острова и морские течения указывали на книгу с картами, даже если ударение не расслышано. Контекст добавляет сведения, но если разные варианты всё ещё подходят, нужно уточнение.

06. Этап 4. Озвучивание музейной подписи 12 минут

ГОВОРИМ УЧЕНИКАМ

Теперь проверим, как написанное превращается в звучащую речь. Синтезатору приходится выбирать ударение и произношение чисел. Слушатель не видит экран. Из-за ошибки он может перепутать предмет или время. Сначала определим нужный смысл, затем прослушаем и проверим результат.

Задание 2. Запишите текст для аудиогида

Карточка школьного музея: «Перед вами атлас. Из этой ткани сшили праздничный костюм. Послушать рассказ экскурсовода можно с 12:05».

1. Определите, что слушатель должен понять о предмете и времени.
2. Напишите вариант для озвучивания: сделайте понятным значение слова «атлас», время запишите словами.
3. Учитель запускает синтез исходного и исправленного текста. Послушайте обе версии.
4. В тетради отметьте фактический результат: значение и ударение; время начала; паузы, которые помогают понять сообщение. Если ошибки не было, запишите это.

Результат: текст для аудиогида и запись проверки. Записывайте то, что услышали.

Для озвучивания учитель использует доступную на своём устройстве функцию синтеза речи с русским голосом.

Для учителя

Ожидаемый смысл: атла́с — ткань; рассказ доступен начиная с двенадцати часов пяти минут. Возможная редакция: «Перед вами атла́с, гладкая ткань. Из неё сшили праздничный костюм. Рассказ экскурсовода можно послушать начиная с двенадцати часов пяти минут». При проверке сверяйте услышанное время с карточкой. Из «можно с 12:05» не следует, что рассказ доступен только в одну минуту.

Кнопка запускает доступный на устройстве русский голос. Если он не найден, используйте знакомый учителю синтезатор. Исправленный текст нужно действительно прослушать: разные голоса по-разному обрабатывают ударение и контекст.

Если синтеза нет, пары обмениваются подготовленными текстами и читают их вслух. Слушатель проверяет, понял ли он предмет и время без взгляда на исходную карточку. Заранее неправильное чтение не задавайте. Это проверка понятности текста человеком; работу синтезатора она не заменяет.

Засчитывайте вывод из прослушивания по трём условиям: значение, время и понятность пауз. Отсутствие ошибок тоже может быть верным результатом. При отсутствии сервиса не отмечайте практику синтеза как проведённую.

ГОВОРИМ УЧЕНИКАМ

В музейной подписи речь о ткани, а двенадцать часов пять минут обозначают начало доступного времени. Эти сведения должны сохраниться в звучании. Запись времени словами и пояснение значения могут помочь, но результат всё равно проверяют на слух. Синтез полезен, когда нужен доступ к написанному без чтения с экрана.

07. Этап 5. Знакомое звучание и проверка отправителя 7 минут

ГОВОРИМ УЧЕНИКАМ

Голосовой чат-бот — программа для разговора с человеком. Она может распознавать речь, выбирать ответ и озвучивать его. Это помогает, например, узнать состояние заказа без ручного поиска. Но разговорное звучание не делает любой ответ правильным, а знакомый тембр сам по себе не подтверждает личность говорящего.

Разберите просьбу

В чате подготовки школьного спектакля появляется голосовое сообщение с нового аккаунта. Голос похож на голос руководителя кружка: «Установите приложение по ссылке, там новая запись репетиции». Школьники раньше получали материалы от руководителя другим способом. Как проверить просьбу до установки?

Для учителя

Подходит: связаться с руководителем через уже известный контакт, попросить подтвердить приложение и способ получения материалов; до проверки ничего не устанавливать. Новый аккаунт и необычный способ требуют проверки. Они ещё не доказывают подделку.

В учебном случае отправитель использовал синтез похожего голоса по прежней записи выступления. По одному сообщению ученики не могут установить это наверняка. Ученики должны выбрать независимый способ проверки, а не угадать подделку на слух.

ГОВОРИМ УЧЕНИКАМ

В этом учебном случае похожий голос создала программа по прежней записи. Школьники не могли доказать это на слух, но могли проверить просьбу через привычный контакт руководителя. Если голосовой сервис просит пароль, код или чужую запись, их нельзя передавать по такому запросу. Для учебных опытов достаточно обезличенных фраз, а чужой голос используют только с разрешения.

Итоговые вопросы

1. Сервис получил запись экскурсии и сделал субтитры. Другой сервис прочитал написанную аннотацию. Как называется каждая задача? Что дано на входе и что получилось?
2. Зачем при обучении распознаванию используют запись вместе с правильным текстом? Почему даже знакомые слова могут быть распознаны неверно в шуме?
3. Чем различаются омофоны, омографы и омонимы? Приведите по одному примеру из урока.
4. Что нужно проверить в озвученной музейной подписи? Почему удачное произношение и знакомый голос не подтверждают, кто отправил сообщение?

Для учителя

1. Распознавание: звуковая запись → текст субтитров. Синтез: текст аннотации → звучащая речь. 2. Пара связывает звучание с нужной записью слов. Шум может скрыть часть звуков, и похожий вариант будет выбран неверно. 3. «Гриб — грипп» или «код — кот»: одинаковое звучание, разные написания. «А́тлас — атла́с»: одинаковое написание, разные ударения и значения. Два значения «косы»: одинаковые написание и звучание, разные значения. 4. Значение ткани, начало доступного времени и понятность пауз. Запись или синтез могут воспроизводить похожий голос; отправителя проверяют отдельно.

ГОВОРИМ УЧЕНИКАМ

Распознавание записывает речь текстом, синтез озвучивает текст. Мы проверяем их результат по исходному сообщению и условиям задачи. Контекст помогает различать слова, а при нехватке сведений нужен вопрос. Сходство голоса не заменяет проверку отправителя.

← Урок 11 · Все уроки 5 класса · Урок 13 →